

M1-E

0400/0420
0400
04-23-01.

0280

OIPE

#5

P.S

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/731,872
DATE: 03/26/2001
TIME: 12:46:57

Input Set : F:\SEQ.LIS\Seq.lis
Output Set: N:\CRF3\03262001\I731872.raw

3 <110> APPLICANT: Dumas Milne Edwards, Jean Baptiste
4 Bougueleret, Lydie
5 Jobert, Severin
7 <120> TITLE OF INVENTION: FULL-LENGTH HUMAN cDNAs ENCODING POTENTIALLY SECRETED PROTEINS
9 <130> FILE REFERENCE: 78.US3.REG
C--> 11 <140> CURRENT APPLICATION NUMBER: US/09/731,872
C--> 11 <141> CURRENT FILING DATE: 2000-12-07
11 <150> PRIOR APPLICATION NUMBER: US 60/169,629
12 <151> PRIOR FILING DATE: 1999-12-08
14 <150> PRIOR APPLICATION NUMBER: US 60/187,470
15 <151> PRIOR FILING DATE: 2000-03-06
17 <160> NUMBER OF SEQ ID NOS: 482
19 <170> SOFTWARE: Patent.pm
21 <210> SEQ ID NO: 1
22 <211> LENGTH: 2201
23 <212> TYPE: DNA
24 <213> ORGANISM: Homo sapiens
26 <220> FEATURE:
27 <221> NAME/KEY: CDS
28 <222> LOCATION: 169..1692
30 <220> FEATURE:
31 <221> NAME/KEY: sig_peptide
32 <222> LOCATION: 169..249
33 <223> OTHER INFORMATION: Von Heijne matrix
34 score 7.15265901862021
35 seq VLLLLLLERGMFS/SP
37 <400> SEQUENCE: 1
38 agatgtgaat agctccacta taccagcctc gtcttccttc cgggggacaa cgtgggtcag 60
39 ggcacagaga gatattttaat gtcaccctct tggggctttc atgggactcc ctctgccaca 120
40 ttttttgag gttgggaaag ttgctagagg cttcagaact ccagccta atg gat ccc 177
41 Met Asp Pro
42 -25
43 aaa ctc ggg aga atg gct gcg tcc ctg ctg gct gtg ctg ctg ctg 225
44 Lys Leu Gly Arg Met Ala Ala Ser Leu Leu Ala Val Leu Leu Leu Leu
45 -20 -15 -10
46 ctg ctg gag cgc ggc atg ttc tcc tca ccc tcc ccg ccc ccg gcg ctg 273
47 Leu Leu Glu Arg Gly Met Phe Ser Ser Pro Ser Pro Pro Pro Ala Leu
48 -5 1 5
49 tta gag aaa gtc ttc cag tac att gac ctc cat cag gat gaa ttt gtg 321
50 Leu Glu Lys Val Phe Gln Tyr Ile Asp Leu His Gln Asp Glu Phe Val
51 10 15 20
52 cag acg ctg aag gag tgg gtg gcc atc gag agc gac tct gtc cag cct 369
53 Gln Thr Leu Lys Glu Trp Val Ala Ile Glu Ser Asp Ser Val Gln Pro
54 25 30 35 40
55 gtg cct cgc ttc aga caa gag ctc ttc aga atg atg gcc gtg gct gcg 417
56 Val Pro Arg Phe Arg Gln Glu Leu Phe Arg Met Met Ala Val Ala Ala
57 45 50 55

ENTERED

RAW SEQUENCE LISTING

DATE: 03/26/2001

PATENT APPLICATION: US/09/731,872

TIME: 12:46:57

Input Set : F:\SEQ.LIS\Seq.lis

Output Set: N:\CRF3\03262001\I731872.raw

58	gac acg ctg cag cgc ctg ggg gcc cgt gtg gcc tcg gtg gac atg ggt	465
59	Asp Thr Leu Gln Arg Leu Gly Ala Arg Val Ala Ser Val Asp Met Gly	
60	60 65 70	
61	cct cag cag ctg ccc gat ggt cag agt ctt cca ata cct ccc gtc atc	513
62	Pro Gln Gln Leu Pro Asp Gly Gln Ser Leu Pro Ile Pro Pro Val Ile	
63	75 80 85	
64	ctg gcc gaa ctg ggg agc gat ccc acg aaa ggc acc gtg tgc ttc tac	561
65	Leu Ala Glu Leu Gly Ser Asp Pro Thr Lys Gly Thr Val Cys Phe Tyr	
66	90 95 100	
67	ggc cac ttg gac gtg cag cct gct gac cgg ggc gat ggg tgg ctc acg	609
68	Gly His Leu Asp Val Gln Pro Ala Asp Arg Gly Asp Gly Trp Leu Thr	
69	105 110 115 120	
70	gac ccc tat gtg ctg acg gag gta gac ggg aaa ctt tat gga cga gga	657
71	Asp Pro Tyr Val Leu Thr Glu Val Asp Gly Lys Leu Tyr Gly Arg Gly	
72	125 130 135	
73	gcg acc gac aac aaa ggc cct gtc ttg gct tgg atc aat gct gtg agc	705
74	Ala Thr Asp Asn Lys Gly Pro Val Leu Ala Trp Ile Asn Ala Val Ser	
75	140 145 150	
76	gcc ttc aga gcc ctg gag caa gat ctt cct gtg aat atc aaa ttc atc	753
77	Ala Phe Arg Ala Leu Glu Gln Asp Leu Pro Val Asn Ile Lys Phe Ile	
78	155 160 165	
79	att gag ggg atg gaa gag gct ggc tct gtt gcc ctg gag gaa ctt gtg	801
80	Ile Glu Gly Met Glu Glu Ala Gly Ser Val Ala Leu Glu Glu Leu Val	
81	170 175 180	
82	gaa aaa gaa aag gac cga ttc ttc tct ggt gtg gac tac att gta att	849
83	Glu Lys Glu Lys Asp Arg Phe Phe Ser Gly Val Asp Tyr Ile Val Ile	
84	185 190 195 200	
85	tca gat aac ctg tgg atc agc caa agg aag cca gca atc act tat gga	897
86	Ser Asp Asn Leu Trp Ile Ser Gln Arg Lys Pro Ala Ile Thr Tyr Gly	
87	205 210 215	
88	acc cgg ggg aac agc tac ttc atg gtg gag gtg aaa tgc aga gac cag	945
89	Thr Arg Gly Asn Ser Tyr Phe Met Val Glu Val Lys Cys Arg Asp Gln	
90	220 225 230	
91	gat ttt cac tca gga acc ttt ggt ggc atc ctt cat gaa cca atg gct	993
92	Asp Phe His Ser Gly Thr Phe Gly Gly Ile Leu His Glu Pro Met Ala	
93	235 240 245	
94	gat ctg gtt gct ctt ctc ggt agc ctg gta gac tcg tct ggt cat atc	1041
95	Asp Leu Val Ala Leu Leu Gly Ser Leu Val Asp Ser Ser Gly His Ile	
96	250 255 260	
97	ctg gtc cct gga atc tat gat gaa gtg gtt cct ctt aca gaa gag gaa	1089
98	Leu Val Pro Gly Ile Tyr Asp Glu Val Val Pro Leu Thr Glu Glu Glu	
99	265 270 275 280	
100	ata aat aca tac aaa gcc atc cat cta gac cta gaa gaa tac cgg aat	1137
101	Ile Asn Thr Tyr Lys Ala Ile His Leu Asp Leu Glu Glu Tyr Arg Asn	
102	285 290 295	
103	agc agc cgg gtt gag aaa ttt ctg ttc gat act aag gag gag att cta	1185
104	Ser Ser Arg Val Glu Lys Phe Leu Phe Asp Thr Lys Glu Glu Ile Leu	
105	300 305 310	
106	atg cac ctc tgg agg tac cca tct ctt tct att cat ggg atc gag ggc	1233

RAW SEQUENCE LISTING DATE: 03/26/2001
 PATENT APPLICATION: US/09/731,872 TIME: 12:46:57

Input Set : F:\SEQ.LIS\Seq.lis
 Output Set: N:\CRF3\03262001\I731872.raw

```

107 Met His Leu Trp Arg Tyr Pro Ser Leu Ser Ile His Gly Ile Glu Gly
108      315      320      325
109 gcg ttt gat gag cct gga act aaa aca gtc ata cct ggc cga gtt ata 1281
110 Ala Phe Asp Glu Pro Gly Thr Lys Thr Val Ile Pro Gly Arg Val Ile
111      330      335      340
112 gga aaa ttt tca atc cgt cta gtc cct cac atg aat gtg tct gcg gtg 1329
113 Gly Lys Phe Ser Ile Arg Leu Val Pro His Met Asn Val Ser Ala Val
114 345      350      355      360
115 gaa aaa cag gtg aca cga cat ctt gaa gat gtg ttc tcc aaa aga aat 1377
116 Glu Lys Gln Val Thr Arg His Leu Glu Asp Val Phe Ser Lys Arg Asn
117      365      370      375
118 agt tcc aac aag atg gtt gtt tcc atg act cta gga cta cac ccg tgg 1425
119 Ser Ser Asn Lys Met Val Val Ser Met Thr Leu Gly Leu His Pro Trp
120      380      385      390
121 att gca aat att gat gac acc cag tat ctc gca gca aaa aga gcg atc 1473
122 Ile Ala Asn Ile Asp Asp Thr Gln Tyr Leu Ala Ala Lys Arg Ala Ile
123      395      400      405
124 aga aca gtg ttt gga aca gaa cca gat atg atc cgg gat gga tcc acc 1521
125 Arg Thr Val Phe Gly Thr Glu Pro Asp Met Ile Arg Asp Gly Ser Thr
126      410      415      420
127 att cca att gcc aaa atg ttc cag gag atc gtc cac aag agc gtg gtg 1569
128 Ile Pro Ile Ala Lys Met Phe Gln Glu Ile Val His Lys Ser Val Val
129 425      430      435      440
130 cta att ccg ctg gga gct gtt gat gat gga gaa cat tcg cag aat gag 1617
131 Leu Ile Pro Leu Gly Ala Val Asp Asp Gly Glu His Ser Gln Asn Glu
132      445      450      455
133 aaa atc aac agg tgg aac tac ata gag gga acc aaa tta ttt gct gcc 1665
134 Lys Ile Asn Arg Trp Asn Tyr Ile Glu Gly Thr Lys Leu Phe Ala Ala
135      460      465      470
136 ttt ttc tta gag atg gcc cag ctc cat taatcacaag aacottctag 1712
137 Phe Phe Leu Glu Met Ala Gln Leu His
138      475      480
139 tctgatctga tccactgaca gattcacctc cccacacatcc ctagacaggg atggaatgta 1772
140 aatatccaga gaatttggtt ctagtatagt acattttccc ttccatttaa aatgtcttgg 1832
141 gatatctgga tcagtaataa aatatttcaa aggcacagat gttggaaatg gtttaagggtc 1892
142 cccactgca cacttcctc aagtcatagc tgcttgagc aacttgattt cccaagtcc 1952
143 tgtgcaatag cccaggatt ggattccttc caaccttta gcatactctc aaccttgcaa 2012
144 tttgattggc ataactcactc cagtttgctt tctaggtcct caagtgtctg tgacacataa 2072
145 tcattccatc caatgatcgc ctttgcttta ccactcttcc cttttatctt attaataaaa 2132
146 atgttggtct ccaccactga aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaagaaaaaa 2192
147 aaaaaaaaaa 2201
149 <210> SEQ ID NO: 2
150 <211> LENGTH: 1631
151 <212> TYPE: DNA
152 <213> ORGANISM: Homo sapiens
154 <220> FEATURE:
155 <221> NAME/KEY: CDS
156 <222> LOCATION: 148..1140
158 <220> FEATURE:

```

RAW SEQUENCE LISTING DATE: 03/26/2001
 PATENT APPLICATION: US/09/731,872 TIME: 12:46:57

Input Set : F:\SEQ.LIS\Seq.lis
 Output Set: N:\CRF3\03262001\I731872.raw

```

159 <221> NAME/KEY: sig_peptide
160 <222> LOCATION: 148..240
161 <223> OTHER INFORMATION: Von Heijne matrix
162     score 10.0910253445132
163     seq LVLLLVTRSPVNA/CL
165 <400> SEQUENCE: 2
166 gtcctgtgcc gccattgtgc ggcgtggc cctcagagg gttcctgtg ctgccgtgc      60
167 cttggacctt cccctcgtct tctcgttcta ctgccccagg agcccggcgg gtccgggact    120
168 cccgtccgtg ccggtgcggg cgccggc atg tgg ctg tgg gag gac cag ggc ggc    174
169                               Met Trp Leu Trp Glu Asp Gln Gly Gly
170                               -30                               -25
171 ctc ctg ggc cct ttc tcc ttc ctg ctg cta gtg ctg ctg ctg gtg acg      222
172 Leu Leu Gly Pro Phe Ser Phe Leu Leu Val Leu Leu Val Thr
173     -20                               -15                               -10
174 cgg agc ccg gtc aat gcc tgc ctc ctc acc ggc agc ctc ttc gtt cta      270
175 Arg Ser Pro Val Asn Ala Cys Leu Leu Thr Gly Ser Leu Phe Val Leu
176     -5                               1                               5                               10
177 ctg cgc gtc ttc agc ttt gag ccg gtg ccc tct tgc agg gcc ctg cag      318
178 Leu Arg Val Phe Ser Phe Glu Pro Val Pro Ser Cys Arg Ala Leu Gln
179     15                               20                               25
180 gtg ctc aag ccc cgg gac cgc att tct gcc atc gcc cac cgt ggc ggc      366
181 Val Leu Lys Pro Arg Asp Arg Ile Ser Ala Ile Ala His Arg Gly Gly
182     30                               35                               40
183 agc cac gac gcg ccc gag aac acg ctg gcg gcc att cgg cag gca gct      414
184 Ser His Asp Ala Pro Glu Asn Thr Leu Ala Ala Ile Arg Gln Ala Ala
185     45                               50                               55
186 aag aat gga gca aca ggc gtg gag ttg gac att gag ttt act tct gac      462
187 Lys Asn Gly Ala Thr Gly Val Glu Leu Asp Ile Glu Phe Thr Ser Asp
188     60                               65                               70
189 ggg att cct gtc tta atg cac gat aac aca gta gat agg acg act gat      510
190 Gly Ile Pro Val Leu Met His Asp Asn Thr Val Asp Arg Thr Thr Asp
191 75                               80                               85                               90
192 ggg act ggg cga ttg tgt gat ttg aca ttt gaa caa att agg aag ctg      558
193 Gly Thr Gly Arg Leu Cys Asp Leu Thr Phe Glu Gln Ile Arg Lys Leu
194     95                               100                               105
195 aat cct gca gca aac cac aga ctc agg aat gat ttc cct gat gaa aag      606
196 Asn Pro Ala Ala Asn His Arg Leu Arg Asn Asp Phe Pro Asp Glu Lys
197     110                               115                               120
198 atc cct acc cta atg gaa gct gtt gca gag tgc cta aac cat aac ctc      654
199 Ile Pro Thr Leu Met Glu Ala Val Ala Glu Cys Leu Asn His Asn Leu
200     125                               130                               135
201 aca atc ttc ttt gat gtc aaa ggc cat gca cac aag gct act gag gct      702
202 Thr Ile Phe Phe Asp Val Lys Gly His Ala His Lys Ala Thr Glu Ala
203     140                               145                               150
204 cta aag aaa atg tat atg gaa ttt cct caa ctg tat aat aat agt gtg      750
205 Leu Lys Lys Met Tyr Met Glu Phe Pro Gln Leu Tyr Asn Asn Ser Val
206 155                               160                               165                               170
207 gtc tgt tct ttc ttg cca gaa gtt atc tac aag atg aga caa aca gat      798
208 Val Cys Ser Phe Leu Pro Glu Val Ile Tyr Lys Met Arg Gln Thr Asp

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/731,872

DATE: 03/26/2001

TIME: 12:46:57

Input Set : F:\SEQ.LIS\Seq.lis

Output Set: N:\CRF3\03262001\I731872.raw

```

209          175          180          185
210 cgg gat gta ata aca gca tta act cac aga cct tgg agc cta agc cat      846
211 Arg Asp Val Ile Thr Ala Leu Thr His Arg Pro Trp Ser Leu Ser His
212          190          195          200
213 aca gga gat ggg aaa cca cgc tat gat act ttc tgg aaa cat ttt ata      894
214 Thr Gly Asp Gly Lys Pro Arg Tyr Asp Thr Phe Trp Lys His Phe Ile
215          205          210          215
216 ttt gtt atg atg gac att ttg ctc gat tgg agc atg cat aat atc ttg      942
217 Phe Val Met Met Asp Ile Leu Leu Asp Trp Ser Met His Asn Ile Leu
218          220          225          230
219 tgg tac ctg tgt gga att tca gct ttc ctc atg caa aag gat ttt gta      990
220 Trp Tyr Leu Cys Gly Ile Ser Ala Phe Leu Met Gln Lys Asp Phe Val
221 235          240          245          250
222 tcc ccg gcc tac ttg aag aag tgg tca gct aaa gga atc cag gtt gtt      1038
223 Ser Pro Ala Tyr Leu Lys Lys Trp Ser Ala Lys Gly Ile Gln Val Val
224          255          260          265
225 ggt tgg act gtt aat acc ttt gat gaa aag agt tac tac gaa tcc cat      1086
226 Gly Trp Thr Val Asn Thr Phe Asp Glu Lys Ser Tyr Tyr Glu Ser His
227          270          275          280
228 ctt ggt tcc agc tat atc act gac agc atg gta gaa gac tgc gaa cct      1134
229 Leu Gly Ser Ser Tyr Ile Thr Asp Ser Met Val Glu Asp Cys Glu Pro
230          285          290          295
231 cac ttc tagactttca cgggtgggacg aaacgggttc agaaactgcc aggggcctca      1190
232 His Phe
233          300
234 tacagggata tcaaaatacc ctttgtgcta gccaggccc tggggaatca ggtgactcac      1250
235 acaaatgcaa tagtttgtca ctgcatTTTT acctgaacca aagctaaacc cgggtgttgcc      1310
236 accatgcacc atggcatgcc agagtcaaac actgttgctc ttgaaaatct ggggtctgaa      1370
237 aaaacgcaca agagccctcg ccctgcccta gctgaggcac acagggagac ccagtgagga      1430
238 taagcacaga ttgaattgta caatttgcag atgcagatgt aaatgcatgg gacatgcatg      1490
239 ataactcaga gttgacattt taaaacttgc cacacttatt tcaaatattt gtactcagct      1550
240 atgttaacat gtactgtaga catcaaaactt gtggccatac taataaaatt attaaaagga      1610
241 gcacaaaaaa aaaaaaaaaa a
242          310
243 <210> SEQ ID NO: 3
244 <211> LENGTH: 1245
245 <212> TYPE: DNA
246 <213> ORGANISM: Homo sapiens
247
248 <220> FEATURE:
249 <221> NAME/KEY: CDS
250 <222> LOCATION: 85..906
251
252 <220> FEATURE:
253 <221> NAME/KEY: sig_peptide
254 <222> LOCATION: 85..135
255 <223> OTHER INFORMATION: Von Heijne matrix
256 score 3.86022363031904
257 seq GFVAALVAGGVAG/VS
258
259 <400> SEQUENCE: 3
260 aaaacatggc ggcgccagc gcgcgaggac gtgatccgct tctgctccgg cttggattgt      60
261 agccttgacg aggtctgagc gacc atg gac cgg cgg ggg ttc gtg gca gcg      111

```

FYI

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

VERIFICATION SUMMARY
PATENT APPLICATION: US/09/731,872

DATE: 03/26/2001
TIME: 12:46:58

Input Set : F:\SEQ.LIS\Seq.lis
Output Set: N:\CRF3\03262001\I731872.raw

L:11 M:270 C: Current Application Number differs, Replaced Current Application No
L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:967 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12
L:1585 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:20
L:3409 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:45
L:6848 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:97
L:7168 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:103
L:7243 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:104
L:7896 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:115
L:8934 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:135
L:8935 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:135
L:9027 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:137
L:9030 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:137
L:10895 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:165
L:12294 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:188
L:12817 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:202
L:13014 M:258 W: Mandatory Feature missing, <223> not found for SEQ ID#:207
L:13014 M:340 W: (46) "n" or "Xaa" used: Feature required, for SEQ ID#:207
L:13192 M:258 W: Mandatory Feature missing, <223> not found for SEQ ID#:210
L:13192 M:340 W: (46) "n" or "Xaa" used: Feature required, for SEQ ID#:210
L:13779 M:258 W: Mandatory Feature missing, <223> not found for SEQ ID#:220
L:13779 M:340 W: (46) "n" or "Xaa" used: Feature required, for SEQ ID#:220
L:15366 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:253
L:16754 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:286
L:18651 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:338
L:18824 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:344
L:19208 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:356
L:19793 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:376
L:19853 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:378
L:19855 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:378
L:21997 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:448
L:22095 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:451
L:22433 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:461